

Результат анализа текста

*The Master and Margarita.txt

Количество
словоупотреблений

> 112 k

*112902

Количество
словоформ

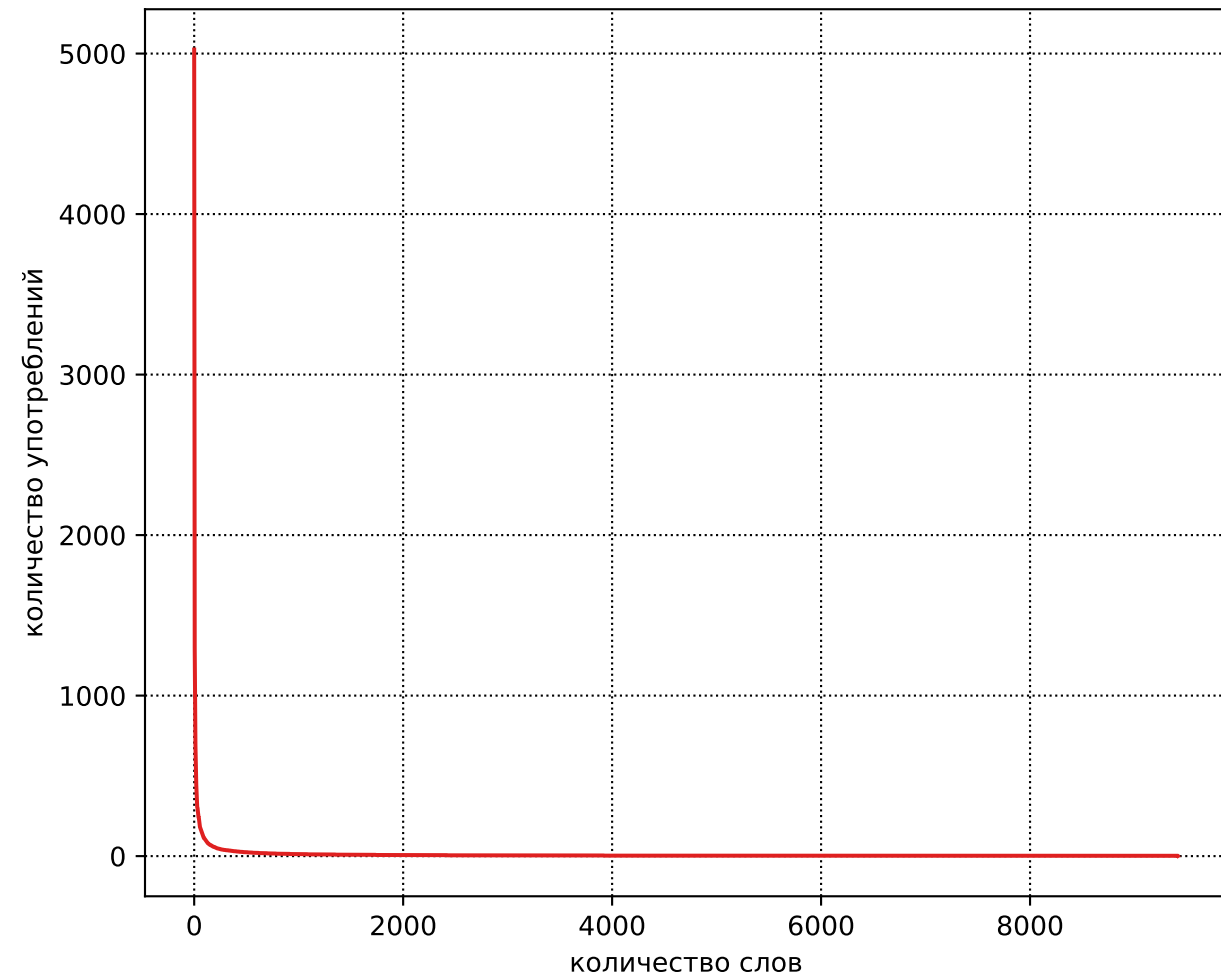
23913

Количество
лемм

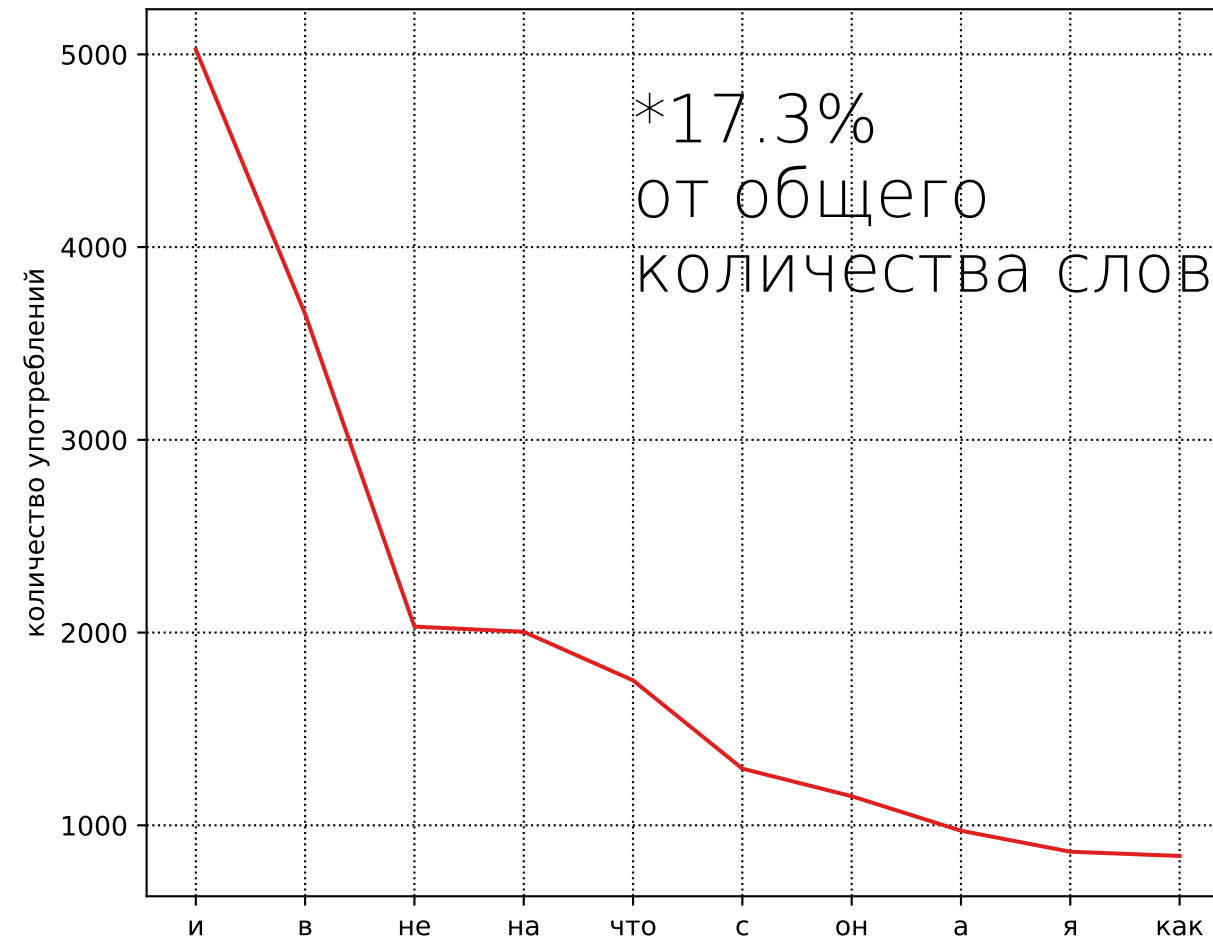
11234

*47.0% от числа
словоформ

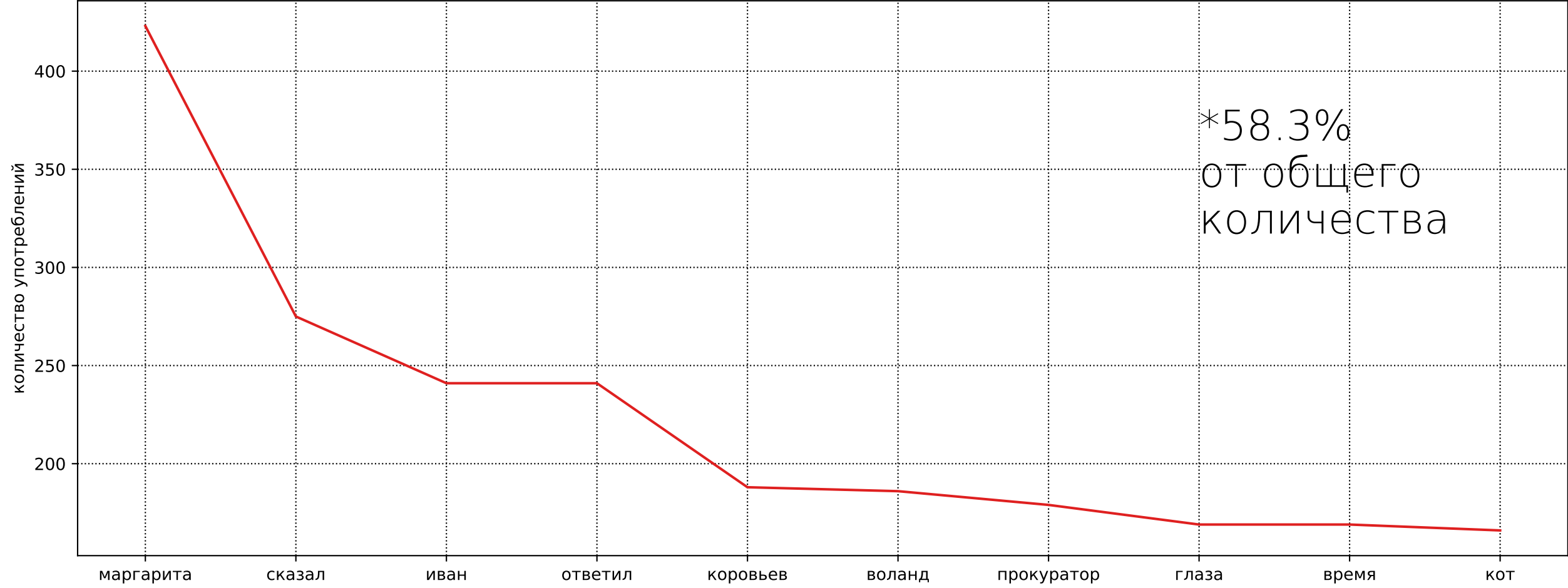
Характер распределения
слов в тексте (max, min)



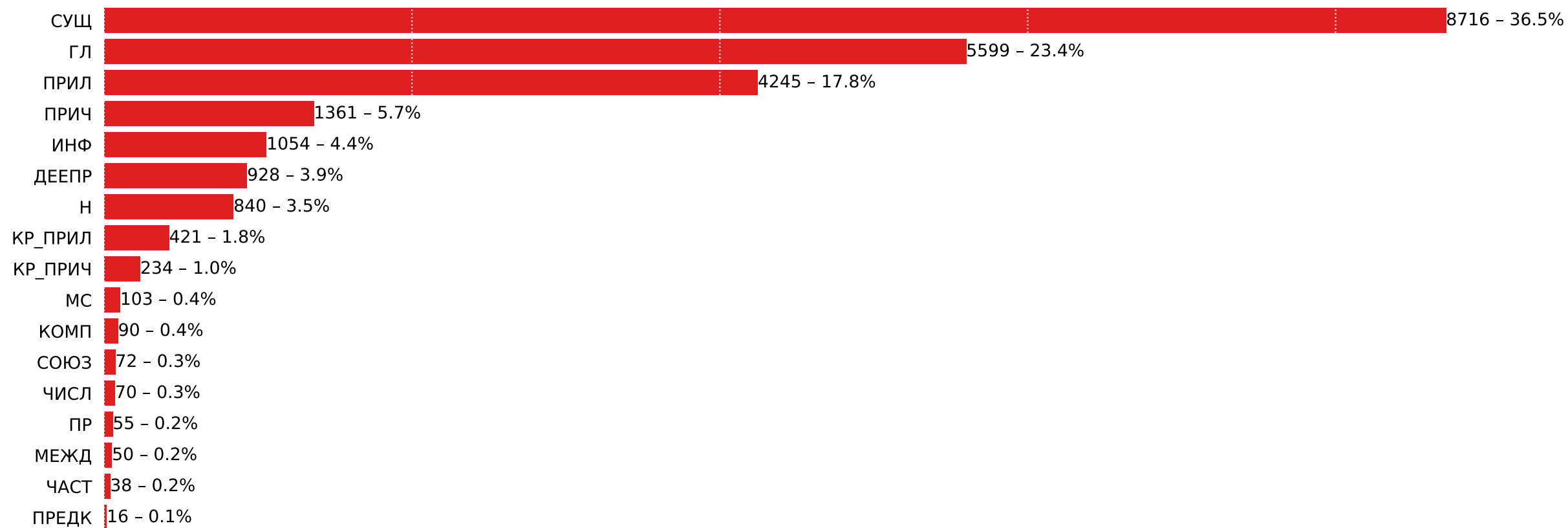
Распределение 10 наиболее частых
слов в тексте



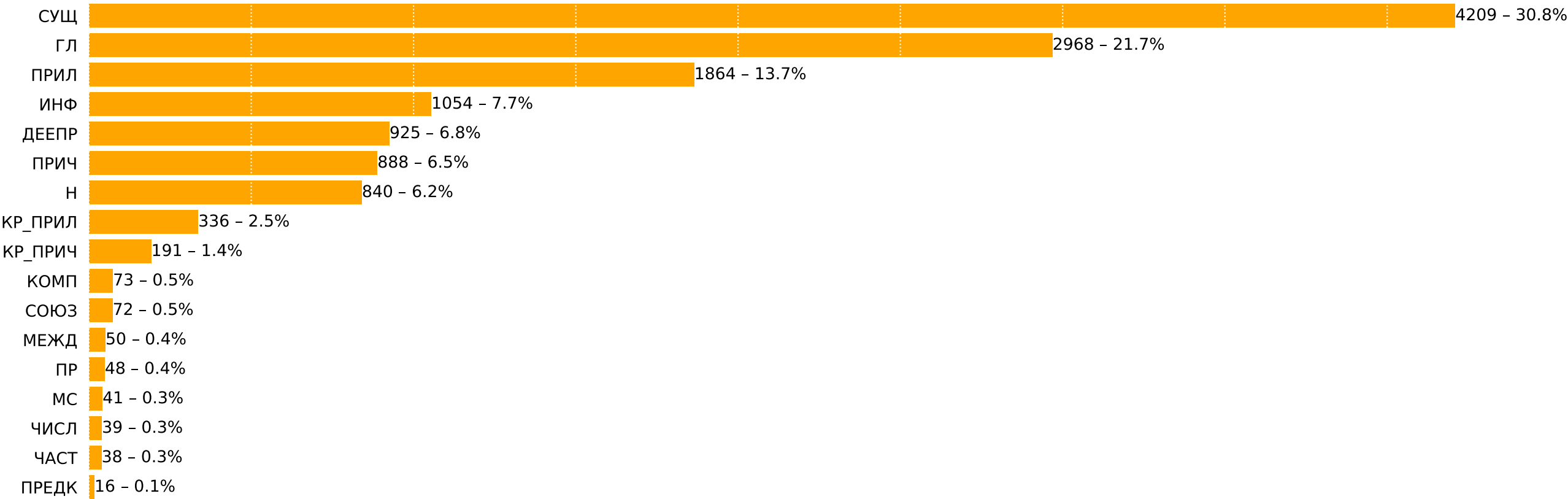
Распределение 10 наиболее частых слов в тексте (без стоп-слов)



Распределение словоформ по частям речи



Распределение лемм
по частям речи



Коэффициент
лексического богатства текста

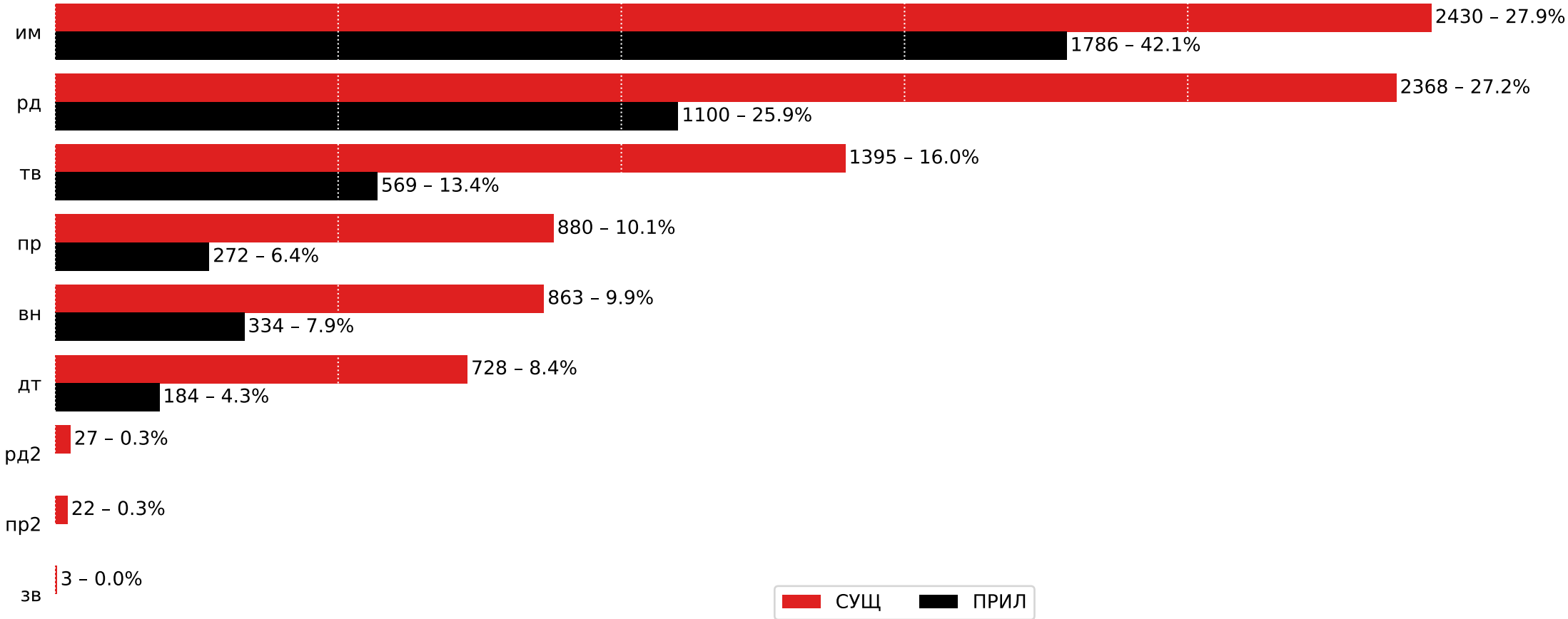
0.47

*отношение числа различных
лемм к общему числу словоформ

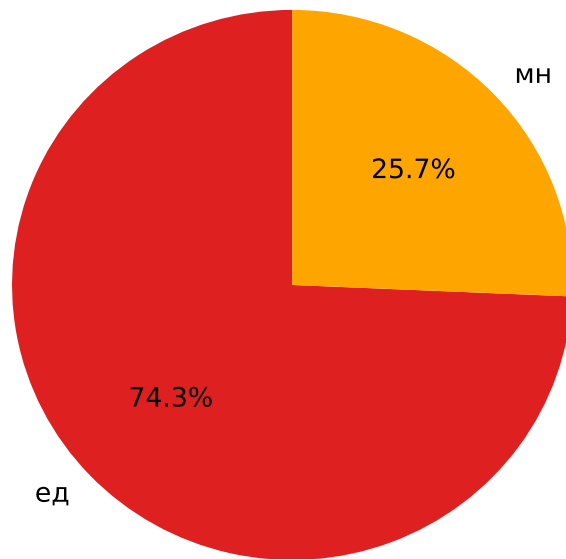
Процент
омонимичных словоформ

45.2%

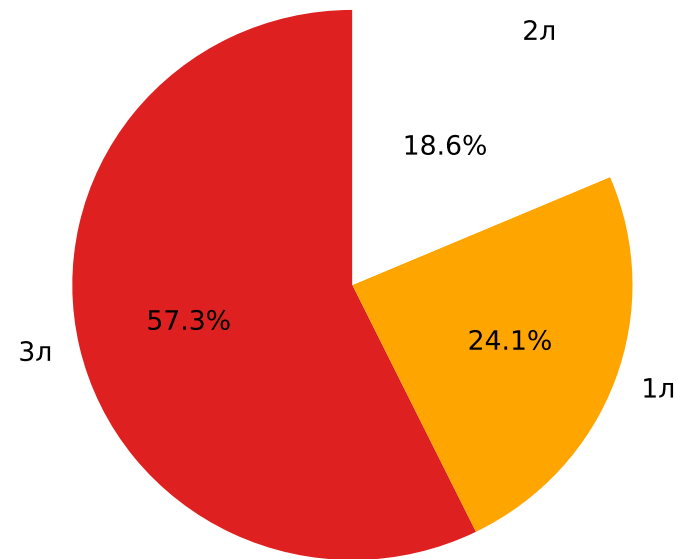
Распределение падежей
существительных и прилагательных



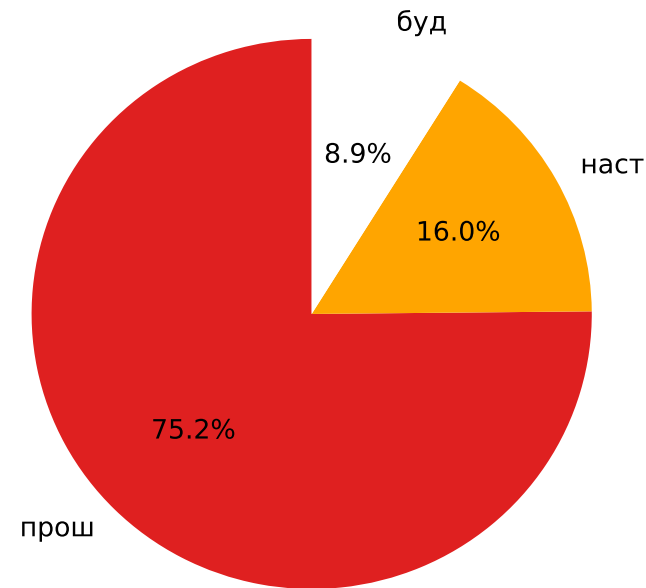
Распределение категории числа
у глаголов



Распределение категории лица
у глаголов



Распределение категории времени
у глаголов



Процент слов,
отсутствующих в словаре

0.8%

*таблица несловарных слов
содержится в файле
./rusult/unknown_words_

ветчинкевич | себастиийских | лиходеевских |
аплодисмент | близехонько | господарские |
елоховской | вороватость | спиридоновке |
утомительнейшее | филиальцев | бумс |
ландшафтик | соболезнующе | тиверия |
пятнажко | лиходеевской | китайцеву |
коровьва | чесунче | чердакчи |
мурлыкнущий | нуар | валерианка |
раздирающе | кентуриях | тысячесвечовыми |
зеленохвостые | выражается | пятидесятилетняя |
хустова | отхохотавшись | намучавшись |
эх-хо-хо | брючонках | перелыгино |
коровьевская | семплеяровской | семплеярову |
гражда | ершалаимском | третьедневочные |
безлесый | волшебств | ершалаимским |
гипнотизерские | бегемотово | пятисвечия |
парчевского | коровьевские | пролежнев |
дипломатически | тацитовых | наказует |
<...>

Средняя оценка точности разбора
каждого слова

0.77

*по данным rymorphy2